

Modèles de régression en présence de compétition

Aurélien Latouche
direction : Sylvie Chevret
co-direction : Raphaël Porcher

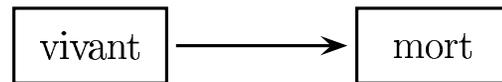
20 Décembre 2004

Plan

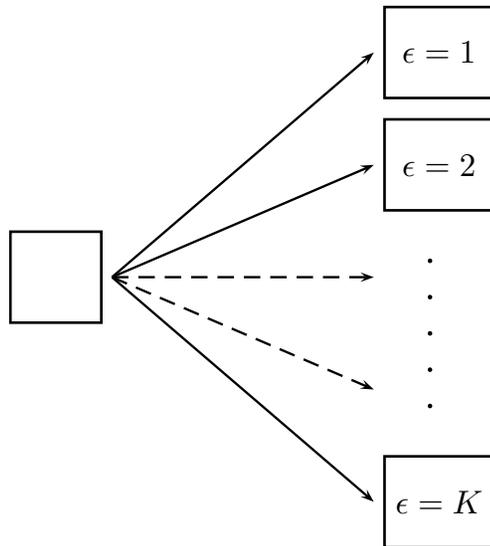
1. Problématique : Risques compétitifs
2. Modélisation par temps latents, identifiabilité
3. Le modèle de Fine & Gray
 - Présentation
 - Inclusion de covariables dépendant du temps
4. Stratégies de modélisation
 - Calcul d'un nombre de sujets nécessaire
 - Modèle mal spécifié
5. Conclusions et Perspectives

1) Problématique

- Analyse de survie : Modèle à 2 états

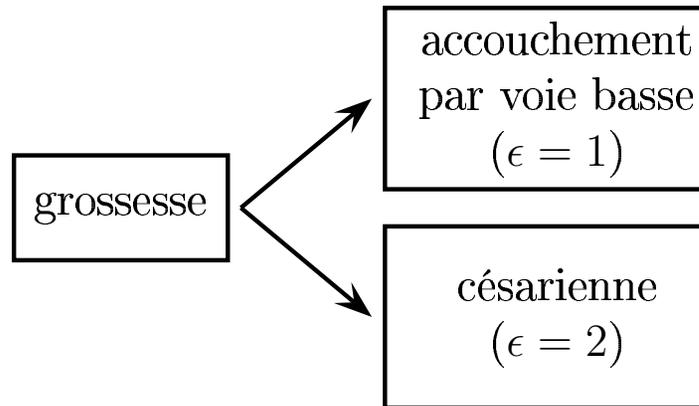


- Compétition : Cas particulier des modèles multi-états

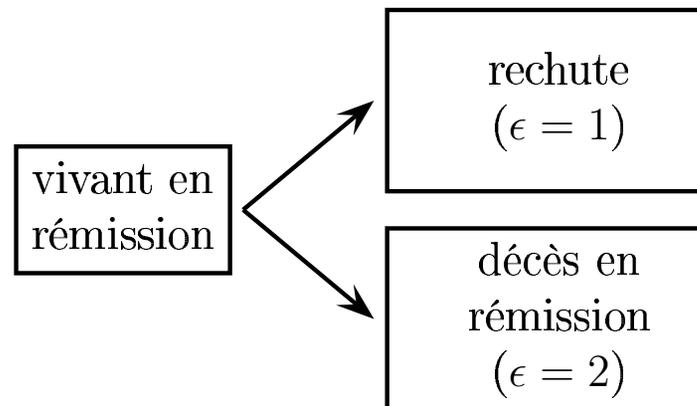


Exemples

Grossesse unique



Hématologie maligne



2) Modélisation par temps latents

- On considère K temps latents
- Les observations consistent en (T, ε) , où
 - $T = \min(T_1, \dots, T_K)$ et
 - $\varepsilon = k$ si $T = T_k$
- Cadre théorique
 - Quantités **nettes** : seule une cause d'intérêt agirait
 - Quantités **brutes** : en présence des K causes

Dans la suite, on considère que $\varepsilon = 1$ est la cause d'intérêt

Quantités brutes

- Fonction de risque instantané cause-spécifique pour l'événement de type 1

$$\lambda_1(t) = \lim_{h \rightarrow 0} \Pr(t \leq T \leq t + h, \varepsilon = 1 | T \geq t) / h$$

- Fonction d'incidence cumulée (fonction de sous-répartition, *subdistribution*) pour l'événement de type 1

$$F_1(t) = \Pr(T \leq t, \varepsilon = 1)$$

Remarque : $\lim_{t \rightarrow \infty} F_1(t) = \Pr(\varepsilon = 1)$

⇒ Contrairement au cas de l'analyse de survie (*i.e.* $K = 1$)

F_1 dépend aussi des λ_k ($k \neq 1$)

Illustration du problème d'identifiabilité ^a

Soit la fonction de survie conjointe

$$H(t_1, t_2) = \exp [1 - a_1 t_1 - a_2 t_2 - \exp\{a_{12}(a_1 t_1 + a_2 t_2)\}]$$

où $a_1, a_2 > 0$ et $a_{12} > -1$

Alors

$$\lambda_k(t) = a_k [1 + a_{12} \exp\{a_{12}(a_1 + a_2)t\}], \quad k = 1, 2$$

Si $a_{12} = 0$, les temps latents sont indépendants

Cependant, soit H^* , définie par

$$H^*(t_1, t_2) = \exp \left[1 - a_1 t_1 - a_2 t_2 - \sum_{k=1}^2 a_k \exp\{a_{12}(a_1 + a_2)t_k\} / (a_1 + a_2) \right]$$

$\forall a_{12}$, les temps latents sont **indépendants** et conduisent aux mêmes λ_k .

^aPrentice et al, 1978

Inférence

- Sous l'hypothèse de temps latents indépendants
 - Inférence possible sur les quantités nettes et brutes
 - Mais ...
 - Hypothèse impossible à tester (Tsiatis, 1975)
 - Problèmes d'identifiabilité
- En absence d'hypothèse d'indépendance des temps latents
 - Seules les quantités brutes sont identifiables
 - Alternative : modéliser λ_1 ou F_1

Modèles de régression

- Risques proportionnels
- Choix de la fonction de risque ?
 - Fonction de risque cause spécifique, λ_1
 - Fonction de risque associée à F_1 , α_1

- Modèle de Cox : modéliser λ_1 conditionnellement à une covariable Z
 - $\lambda_1(t|Z) = \lambda_0(t) \exp(\beta Z)$

3) Le modèle de Fine & Gray

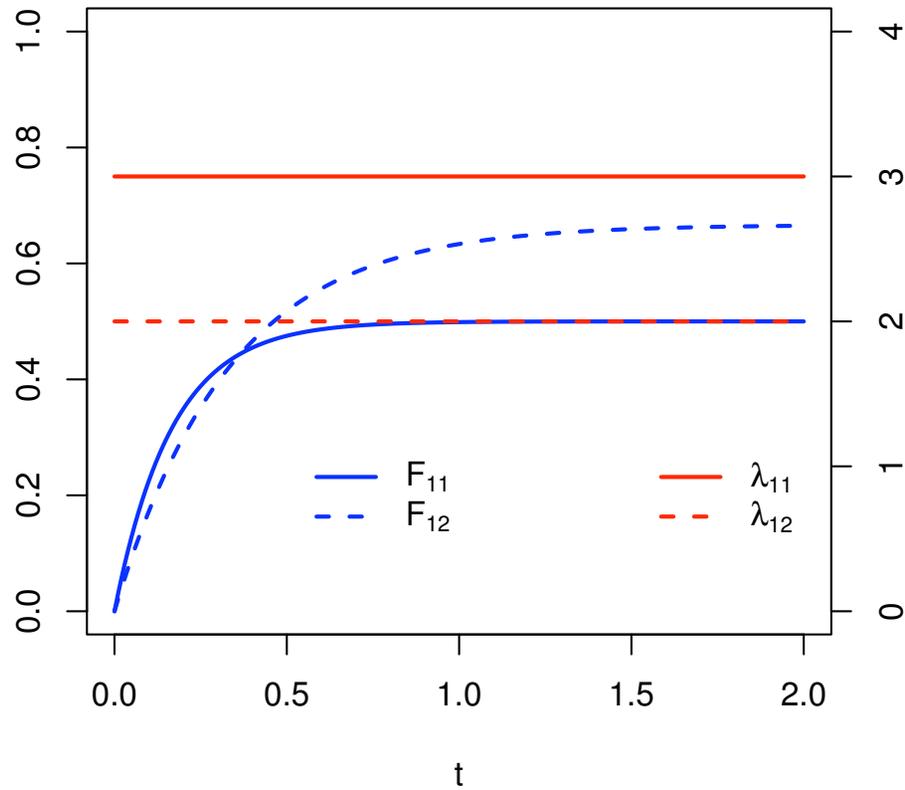
La fonction de risque associée à la fonction d'incidence cumulée (ou fonction de risque de sous-répartition (*subdistribution hazard*))

$$\alpha_1(t) = -\frac{d}{dt} \log\{1 - F_1(t)\}$$

- L'effet d'une covariable sur cette fonction de risque se traduit directement sur la fonction de sous-répartition

Effets des covariables sur les deux fonctions

- $\lambda_{11} = 3$ et $\lambda_{12} = 2$
- $\lambda_{21} = 3$ et $\lambda_{22} = 1$



Le modèle de Fine & Gray (2)

$$\alpha_1(t) = \lim_{h \rightarrow 0} \frac{\Pr\{t \leq T \leq t + h, \varepsilon = 1 | (T \geq t) \cup (T \leq t \cap \varepsilon \neq 1)\}}{h}$$

- C'est la fonction de risque instantané de la pseudo-variable aléatoire

$$T^* = 1_{[\varepsilon=1]} \times T + 1_{[\varepsilon \neq 1]} \times \infty$$

Le modèle de Fine & Gray (3)

- Le modèle à risques proportionnels s'écrit

$$\alpha_1(t; Z) = \alpha_0(t) \exp(\gamma Z)$$

où $\exp(\gamma) = \theta$, dit *Subdistribution hazard ratio*

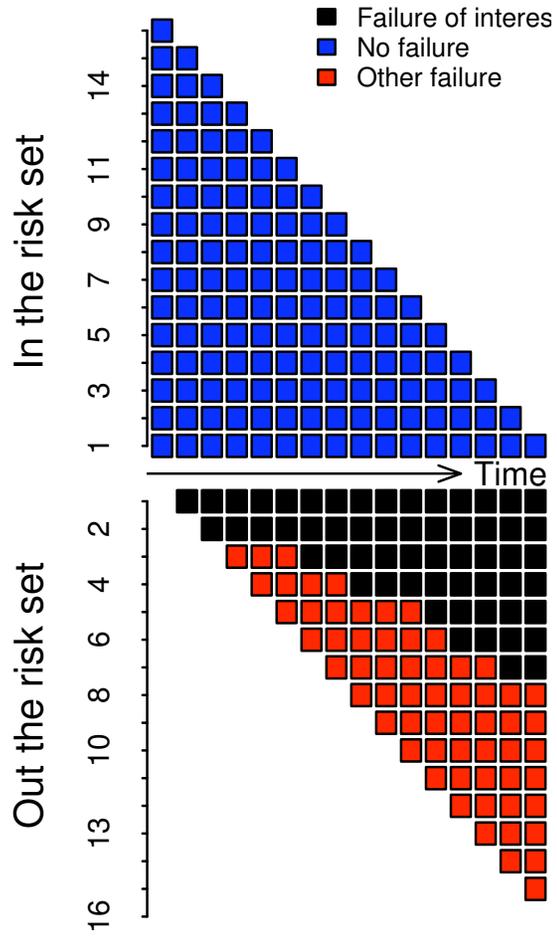
- Propriétés asymptotiques de l'estimateur de γ :
 - En l'absence de censure ou présence de censure à droite administrative : se déduisent de celles d'un modèle de Cox (vraisemblance partielle)
 - Dans les autres situations de censure à droite : Estimateur pondéré (*Inverse Probability Weighting*)

- En l'absence de censure, on définit

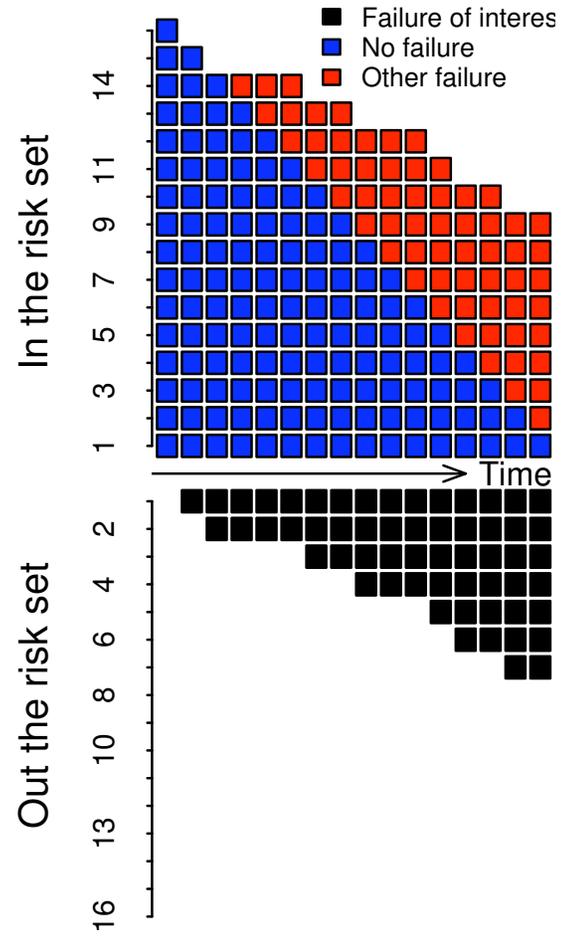
$$\mathcal{R}_i = \{j : (T_i < T_j) \cup (T_i \geq T_j \cap \varepsilon_j \neq 1)\}$$

- Il se compose à l'instant $t = T_i$ (date d'événement d'intérêt du sujet i)
 1. Individus j n'ayant expérimenté aucun événement avant t
(idem Cox)
 2. Individus j ayant expérimenté un événement en compétition avant t

Cox model



Fine and Gray model



Inclusion de covariables dépendant du temps

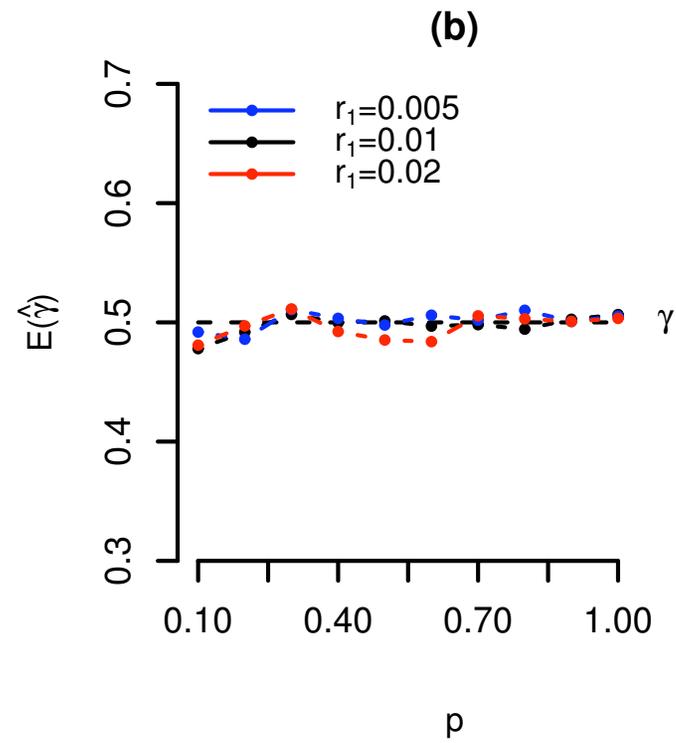
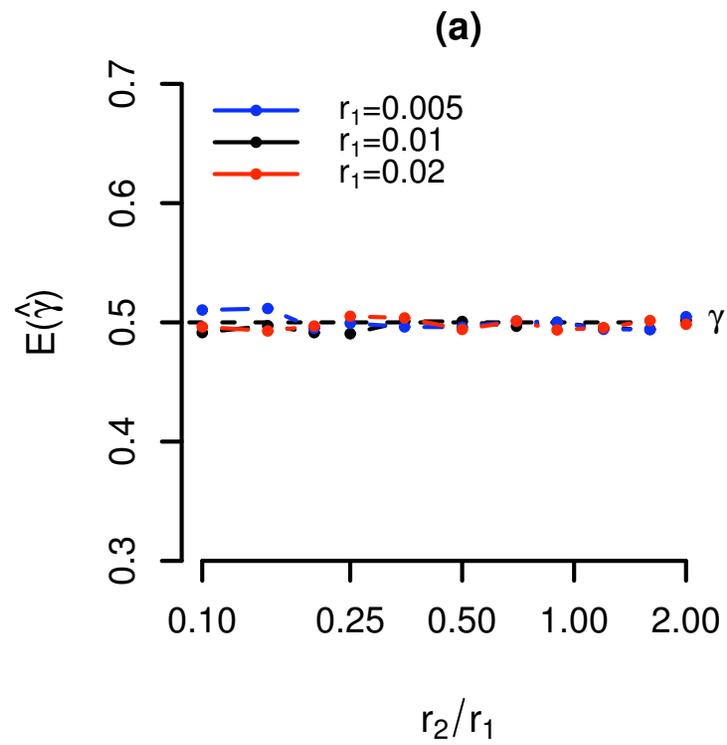
- Exemples nombreux en recherche clinique
 - Données longitudinales (CD4, taux d'immunoglobuline)
 - Covariables binaires, souvent indicatrices de survenue d'un événement (survenue de GvHD après allogreffe de moelle)
- Problème : Observation de $X(t)$ peut être empêchée par la survenue d'un événement en compétition (par exemple si absorbant)

“A note on including time-dependent covariate in regression model for competing risks data”, A. Latouche et al, soumis

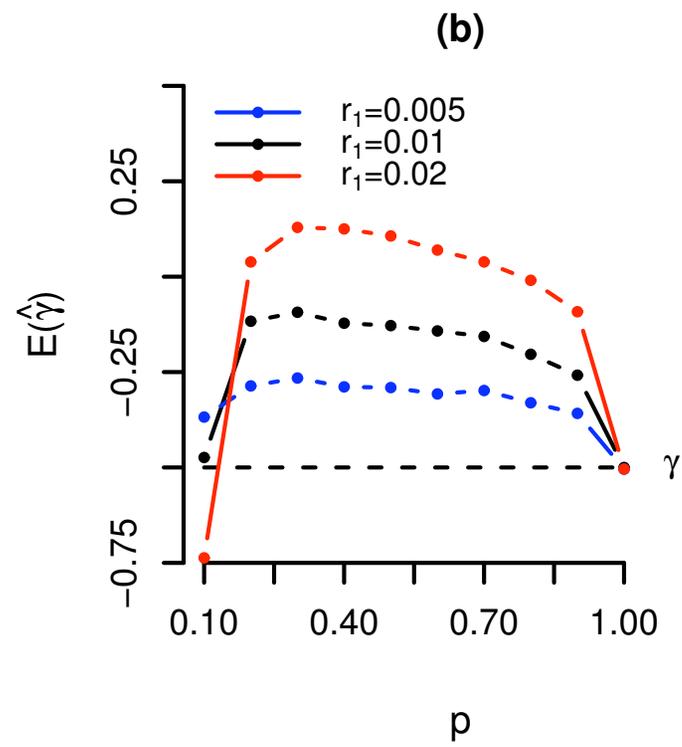
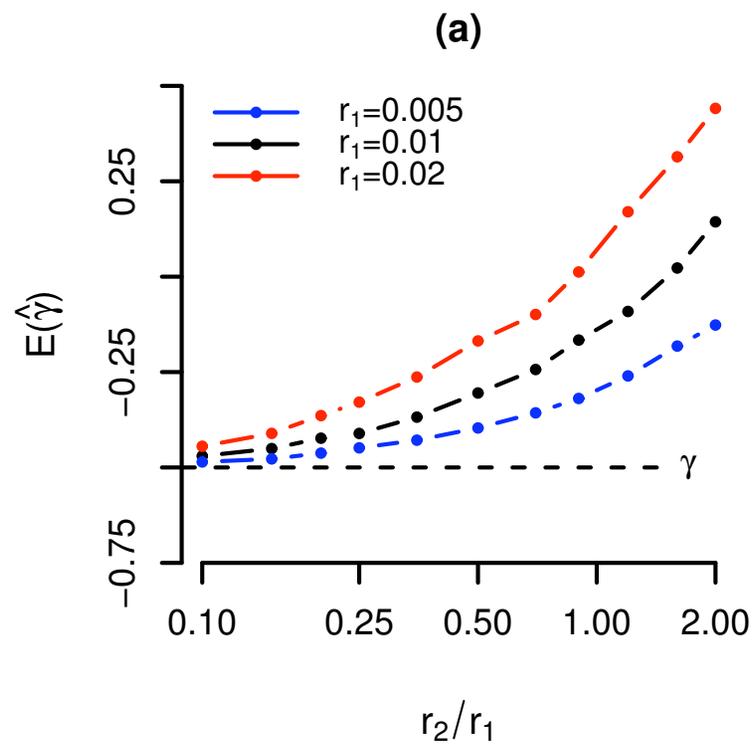
Exemple

- 2 risques en compétition : rechute et décès en rémission
 - **Objectif** : Estimer l'effet de la survenue de aGvHD sur la rechute
 - Introduction de la covariable $X(t) = \mathbf{1}_{\{T_{\text{aGvHD}} \leq t\}}$
- Analyse naïve : effet significatif pour F & G , non significatif pour Cox
- Etude de simulation : biais dans l'estimation de l'effet γ de $X(t)$

$X(t)$ observable



$X(t)$ non observable



4) Stratégies de modélisation

- En pratique, que faire ?
- Opposer ces modèles ?
- Quantités d'intérêt différentes et/mais complémentaires

Calcul d'un nombre de sujets nécessaire

- En présence de compétition, développer une formule pour des covariables binaires potentiellement corrélées
- Extension des formules Schoenfeld (86) et Schmoor (99) dans le cadre du modèle de Cox en absence de compétition

On considère

$$\alpha_1(t; X, Y) = \alpha_0(t) \exp(\gamma X + bY)$$

- X et Y binaires (éventuellement corrélées, ρ)
- $H_0 : \{\gamma = 0\}$
- Statistique (partielle) de Wald : $\sqrt{n}\hat{\gamma}$

“Sample size formula for proportional hazards modelling of competing risks”,
A. Latouche et al, *Statistics in Medicine*, 2004, **23** : 3263-3274

Nombre de sujets nécessaire

$$n = \frac{e}{\psi} = \frac{(u_{\alpha/2} + u_{\beta})^2}{(\log \theta)^2 p (1 - p) \psi (1 - \rho^2)}$$

- $\theta = \exp(\gamma)$
- $p = \Pr(X = 1)$
- ψ , proportion d'événements d'intérêt à la date d'analyse T_a
- α et β , risques d'erreur de type I et II

Expression similaire aux formules de Schoenfeld et Schmoor

– quantités identiques $(\psi, p, \rho, \alpha, \beta)$

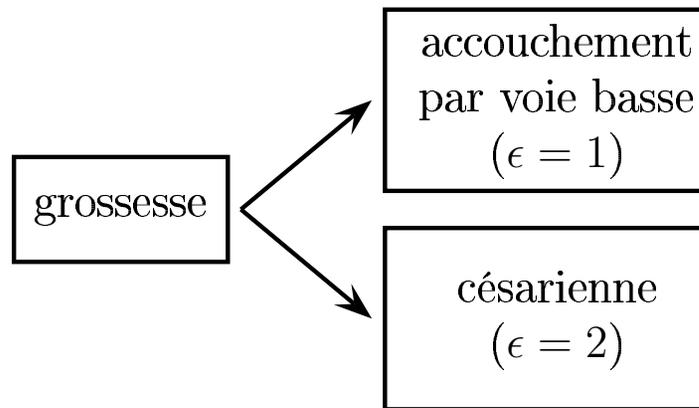
– mais ... rapport θ de fonctions de risque différentes

$$\theta = \frac{\log\{1 - F_1(t; X = 1, Y)\}}{\log\{1 - F_1(t; X = 0, Y)\}}$$

s'obtient directement à partir d'incidences cumulées

Exemple : (re)Planification d'un essai évaluant l'efficacité d'un inducteur du travail

- Essai randomisé en double aveugle
- Critère de jugement : délai d'accouchement par voie basse



Calcul du nombre de sujets

T_a	\hat{F}_1	S	E	$\theta = \frac{\log(1-\hat{F}_E(T_a))}{\log(1-\hat{F}_S(T_a))}$	e	$\psi(\%)$	n
12H	$1 - \widehat{KM}$	0.20	0.30	1.6	191	25.0%	218
	\widehat{CIF}	0.20	0.30	1.6	191		218
24H	$1 - \widehat{KM}$	0.60	0.75	1.5	245	63.5%	364
	\widehat{CIF}	0.60	0.67	1.2	1160		1596
48H	$1 - \widehat{KM}$	0.865	0.90	1.2	1357	77.5%	1546
	\widehat{CIF}	0.77	0.78	1.03	47339		70131

Modèle mal spécifié

- A ce stade, meilleure connaissance du modèle de F & G
- Choix d'un des modèles pour l'analyse
- Approche naturelle (physiologique) : modification du risque cause-spécifique

⇒ Qu'estime le modèle de Fine & Gray si l'on suppose que les données sont issues d'un modèle de Cox ?

“Misspecified regression model for the cumulative incidence function”, A .
Latouche et al, soumis

Hypothèses

- 2 risques en compétition (T_1, T_2)
- Modèle paramétrique pour $(T_1, T_2) \sim \text{ACBVE}$
- Paramètres ACBVE choisis pour vérifier

$$(\mathcal{M}) \begin{cases} \lambda_1(t; Z) = \lambda_{10}(t) \exp(\beta_1 Z) \\ \lambda_2(t; Z) = \lambda_{20}(t) \exp(\beta_2 Z) \end{cases}$$

- Estimation de l'effet de Z dans un modèle de Fine et Gray

$$\alpha_1(t; Z) = \alpha_{10}(t) \exp(\gamma Z)$$

Expression de γ

$$\begin{aligned} \gamma(t) &= \beta_1 + [a - \Psi(\beta_1, \beta_2)]t \\ &+ \log \left[\frac{\Psi(\beta_1, \beta_2)}{a} \times \frac{a_2 + a_1 \exp(-at)}{a_2 \exp(\beta_2) + a_1 \exp(\beta_1) \exp[-\Psi(\beta_1, \beta_2)t]} \right] \end{aligned}$$

γ est donc une fonction

1. compliquée!
2. sens de variation non-trivial
3. dépendant des coefficients de régression pour les 2 risques en compétition

Estimateur de γ

L'estimateur du maximum de vraisemblance partielle, $\hat{\gamma}$, de γ est un estimateur convergent de γ^* où γ^* est la solution de

$$\int_0^\infty E \left\{ f_1(t; Z) \left[Z - \frac{E[Z \exp(\gamma Z)(1 - F_1(t; Z))]}{E[\exp(\gamma Z)(1 - F_1(t; Z))]} \right] \right\} dt = 0.$$

Sous certaines hypothèses on montre que

$$\boxed{\gamma^* \simeq c_0 + c_1 \beta_1}$$

où $c_0 = c_0(\beta_2, \text{Var}(Z))$ et $c_1 = c_1(\beta_2, E(Z))$

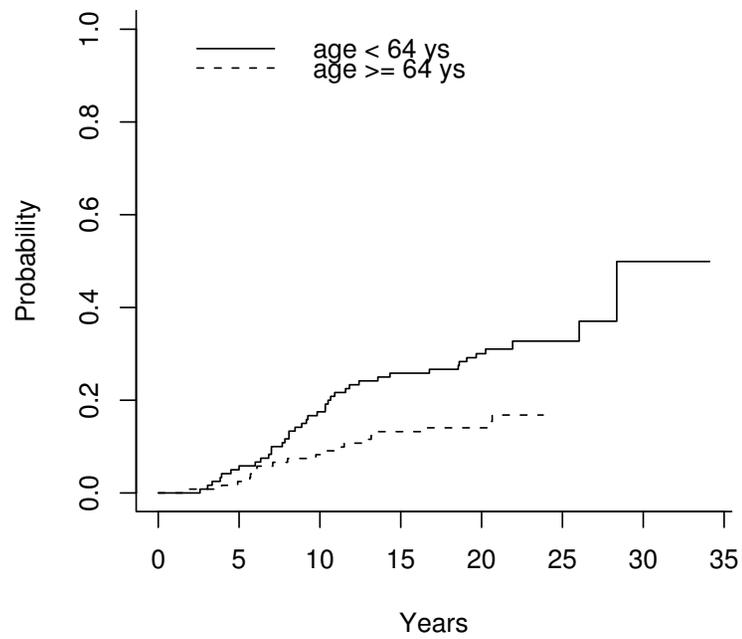
Simulations (1)^a

a_1	a_2	a_{12}	β_1	β_2	γ^*	mean($\hat{\gamma}$)(<i>se</i>)
0.5	1	0	0.1	0.5	-0.296	-0.301 (0.186)
0.5	1	0	0.2	0.5	-0.205	-0.213 (0.181)
0.5	1	0	0.3	0.5	-0.114	-0.128 (0.183)
0.5	1	0	0.4	0.5	-0.022	-0.038 (0.177)
0.5	1	0	0.5	0.5	0.068	0.049 (0.174)
0.5	1	0.5	0.1	0.5	-0.296	-0.308 (0.188)
0.5	1	1	0.1	0.5	-0.296	-0.306 (0.189)
0.5	1	1.5	0.1	0.5	-0.296	-0.307 (0.189)
0.5	1	2	0.1	0.5	-0.296	-0.300 (0.189)
0.5	1	2.5	0.1	0.5	-0.296	-0.304 (0.186)

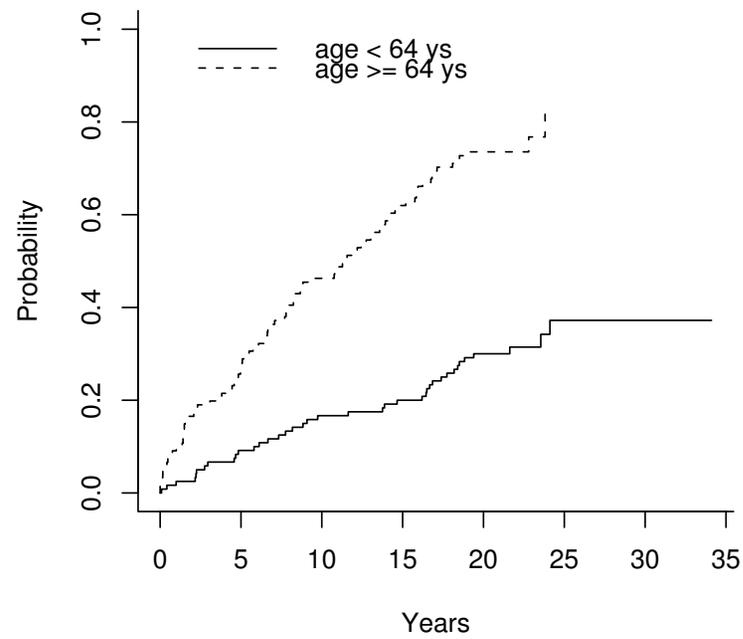
^a5000 \times ($n = 400$)

mgus (Therneau et Grambsch, 2000)

Plasma cell malignancy



Death



mgus

Effect of age	Estimated regression parameter (Standard Error)	
Failure cause	Cause-specific $\hat{\beta}_i$	Subdistribution $\hat{\gamma}_i$
Plasma cell malignancy	-0.225 (0.285)	-0.842 (0.278)
Death as first event	1.39 (0.198)	1.31 (0.192)

$\gamma_1^* = -1.10$, en considérant que $(a_1, a_2) = (0.0198, 0.0405)$

5) Conclusion et Perspectives

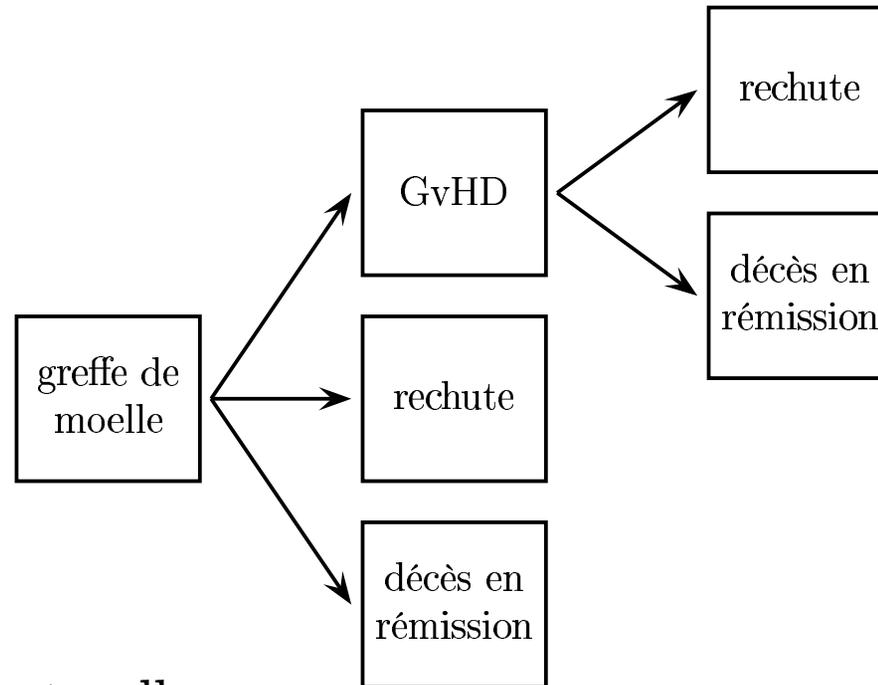
1. Si l'intérêt premier réside dans l'effet d'une exposition sur l'incidence cumulée de l'événement
 - Intérêt en Santé publique (études épidémiologiques, médico-économiques, ...)
 - Choix : Fonctions de risque associées à la fonction de sous-répartition (Modèle de Fine and Gray)
 - Limites : Interprétation conjointe, introduction de covariables dépendant du temps ?
2. Si l'intérêt premier réside dans l'effet d'une exposition sur le risque instantané de l'événement
 - Intérêt en recherche clinique (études thérapeutiques et pronostiques)
 - Choix : Fonctions de risque cause-spécifique (Modèle de Cox)
 - Limites : Pas d'interprétation secondaire en termes de probabilités d'événement

Perspectives

- Risques compétitifs : Modèle de régression pour la probabilité conditionnelle

$$CP_1(t) = \Pr(\text{Evénement d'intérêt avant } t | \text{Pas d'autre événement avant } t)$$

– Risques semi-compétitifs



- Extension naturelle
 - Le décès empêche la rechute ou la GvHD
 - La rechute empêche la GvHD
- $\Rightarrow T_{\text{GvHD}} \leq T_{\text{Rechute}} \leq T_{\text{Décès}}$